数理解析研究所講究録 1375 巻 2004 年 214-220

Mining Indirect Association Rules

大阪府立大学理学系研究科 濱野 慎一* (Shinichi Hamano)
大阪府立大学総合科学部 向内 康人** (Yasuhito Mukouchi)
大阪府立大学総合科学部 佐藤 優子** (Masako Sato)
*Graduate School of Sciences, Osaka Prefecture University
**College of Integrated Arts and Sciences, Osaka Prefecture University

Abstract. A large database, such as POS data, could give us many insights about customer behavior. Many techniques and measures have been proposed to extract the interesting rules. As the study of Association rule mining has proceeded, the rules about items that are not bought together at the same transaction have been regarded as important. Although this concept, Negative Association rule mining, is quite useful, it is difficult for the user to analyze the interestingness of Negative Association rules because we would get them too many. To settle this issue, Indirect Association rule mining has proposed. This is an effective way not only to analyze customer behavior but also to understand the situations of competing. In this paper, we propose a new framework of Indirect Association rule via mediator and a new measure μ based on measures P_A and P_D due to Zhang to mine Negative Association rules effectively without the domain knowledge. Our measure μ has the advantage over the measure IS that is proposed with the first framework of Indirect Association rule mining, and satisfies all of the well-known properties for a good measure. Finally, we are going to analyze the retail data and present interpretations for derived Indirect Association rules.

1 Introduction

Past a decade, Association rule mining proposed in [1] has attracted a great deal of attention. This is now an important tool in CRM (*Customer Relationship Management*) for targeting customers, enhancing cross-selling, and identifying trends and needs. The most fundamental measure, support, has been used to derive the frequent item-sets that have the support above a user-specified threshold, min-support. However, there exist many unconsidered item-sets that could give us great insights into business data. This concept derives the Negative Association rule mining [6]. A Negative Association rule such that there is a negative correlation between the item-sets X and Y enables us to figure out behavioral pattern of customers. It is impossible to analyze every single Negative Association rule in detail because the number of derived Negative Association rules could be large. In order to pick out the valid Negative Associations, Indirect Association rule mining was introduced by Tan et al. [13]. Consider an item-pair $\{a, b\}$ with a low support value, i.e., that are not bought together at the same transaction. If there is an item-set M such that the presence of a and b are highly dependent on items in M, then the pair $\{a, b\}$ are said to be indirectly associated via M. They have called such an item-set M as a *mediator* for $\{a, b\}$, and adapted measure IS to represent dependency between mediator M and each item $x \in \{a, b\}$ [9–12]. IS measure is deeply related with both measures of Interest [4] and Support. Indirect association above, however, does not reflect correlations between a mediator and items in the pair, and moreover, IS is a symmetric measure with no direction. In this paper, we propose a new framework of Indirect Association rule via mediator to enable us to analyze targeting customers, competitors, potential customers, and so on. For an item x, the notation $\neg x$ means the event that a transaction does not contain the item x. Indirect Association rule in our framework is of the form $(M \Rightarrow \beta_a, M \Rightarrow \beta_b)$ for a rare item-pair $\{a, b\}$ and mediator M, where $\beta_x \in \{x, \neg x\}$ for each $x \in \{a, b\}$. A rule $M \Rightarrow \beta_x$ implies a positive correlation between M and β_x, and thus $M \Rightarrow \neg x$ is a Negative Association rule such that there is a Negative correlation between M and x. Mediator M is common item-sets for the customer who is going to buy either a or b. These rules are effective for the users to know the similarity or difference about the behavioral pattern of customers. As a measure for an implication rule $M \Rightarrow \beta_x$, we introduce a new measure μ with direction. Measure μ is based on measures P_A and P_D proposed by Zhang [15] and is shown to mine both of Association rules and Negative Association rules effectively without the domain knowledge. The measures P_A and P_D were presented for a measure of association and disassociation

(Negative Association) relationships between item-sets, respectively, and satisfies all of the well-known properties for a good measure[15]. In the next section, we present our framework of Indirect Association rule via mediator. In section 3, we define measure μ in more general form and investigate properties of measure μ. Moreover, we compare it with various measures such as IS, confidence and so on. Finally, we are going to experiment the real retail data and present interpretations for derived Indirect Association rules.

2 Indirect Association rules

Let I be the finite set of items, and D be the set of customer transactions, called ***database***, where each transaction T is a set of items such that $T \subseteq I$. We denote items of I by $a, b, c, \cdots$ and subsets of I, called *item-sets*, by X, Y, M and so on. In this paper, an item-set X means not only the subset of I but also the event that a transaction contains all items in the set X, and $P(X)$ denotes the probability that a transaction contains the set X. Moreover, by $\neg X$, we denote the negation of the event X, i.e., a transaction does not contain at least one item in X, and thus $\neg\neg X = X$ and $P(\neg X) = 1 - P(X)$. Let us define the following sets:

$$I_f = \{a \in I \mid P(a) \geq t_f\}, \qquad RP = \{\{a,b\} \mid a, b \in I_f,\ P(a,b) \leq t_r\},$$

where t_f and t_r are a ***frequent item threshold*** and a ***rare item-pair threshold*** specified by the user, respectively, such that $t_r < t_f$. Note that if $t_r < t_f^2$, then there is a negative correlation between two events a and b because of $P(a,b) \leq t_r < t_f^2 \leq P(a)P(b)$ for $a, b \in I_f$. An item $a \in I_f$ is called a ***frequent*** item. Note that items in pairs contained in RP are always frequent items. A pair $\{a,b\} \in RP$, called a ***rare item-pair***, is rarely present together in the same transaction, and thus has been treated as uninteresting item-set because of the absence of statistical significance. However, we believe that rare item-pairs could give us great insights by deriving with a common item-sets mediator as shown below. A ***mediator*** M for a rare item-pair $\{a,b\}$ is an item-set of I disjointed to $\{a,b\}$. If there is a mediator M that is highly dependent with both a and b, then such a pair is expected to relate to each other via M. The mediator helps to improve the interpretability of the extracted item-sets by identifying the context for which the Negative Association between a and b is interesting. Let $x \in \{a,b\}$. We call a rule of the form $M \Rightarrow x$ a (positive) Association rule, and $M \Rightarrow \neg x$ a Negative Association rule. A positive (resp., Negative) Association rule represents that there is a positive (resp., negative) correlation between M and x. A Negative Association rule figures out that customers who are going to buy M tend not to buy x in the same transaction. Clearly there is a positive correlation between M and x iff $P(M \mid x) > P(M \mid \neg x)$ iff there is a negative correlation between M and $\neg x$.

Let the notation β_x denote x if $P(M \mid x) \geq P(M \mid \neg x)$, otherwise let β_x denote $\neg x$. Therefore, in either case, $P(M \mid \beta_x) \geq P(M \mid \neg\beta_x)$ holds. As a measure for an implication rule $M \Rightarrow \beta_x$, we define a new measure μ as follows:

$$\mu(M \Rightarrow \beta_x) = \frac{P(M \mid \beta_x) - P(M \mid \neg\beta_x)}{P(M \mid \beta_x)} = \begin{cases} \dfrac{P(M \wedge x) - P(M)P(x)}{P(M \wedge x)(1 - P(x))}, & \text{if } \beta_x = x, \\[2ex] \dfrac{P(M)P(x) - P(M \mid \neg x)}{P(x)(P(M) - P(M \wedge x))}, & \text{if } \beta_x = \neg x, \end{cases}$$

The numerator denotes a conditional measure of a relationship of M with β_x, and the denominator makes it normalized. Hence unlike confidence, the measure μ is normalized like the statistical notion of correlation. Furthermore, the measure μ is directional and measures actual implication as opposed to co-occurrence. By the definition of μ, it follows that :

(i) $0 < \mu(M \Rightarrow \beta_x) \leq 1$, if there is a positive correlation between M and β_x,
(ii) $\mu(M \Rightarrow \beta_x) = 0$, if M and x are statistically independent,
(iii) $\mu(M \Rightarrow \beta_x) = 1$, if M is completely dependent to x, i.e., $P(M \mid \neg\beta_x) = 0$.

Let us define a new framework of Indirect Association via mediator.

Definition 1. *Let $\{a,b\} \in RP$ be a rare item-pair, and let M be a mediator of $\{a,b\}$, i.e., $M \cap \{a,b\} = \phi$. A pair of association rules $(M \Rightarrow \beta_a, M \Rightarrow \beta_b)$ is an Indirect Association rule via a mediator M, denoted by $(M; \{\beta_a, \beta_b\})$, if*
(1) $P(M \wedge \beta_x) \geq t_m$, (Mediator Support Condition),
(2) $\mu(M \Rightarrow \beta_x) \geq t_\mu$, $x \in \{a,b\}$, (Mediator Dependence Condition).
And the rare item-pair $\{a,b\}$ is indirectly associated via mediator M.

Mediator Support Condition (1) is for the statistical significance between the mediator and the pair $\{\beta_a, \beta_b\}$. Marketers regard statistical significance as important because they would like to know the common behavior of customers. That is, a mediator stands for the basic behavior of customers and relations between a mediator and each of rare item-pairs stand for characteristic behavior of customers. Therefore this condition is necessary as statistical significance. Mediator Dependence Condition (2) is for measuring the interestingness of the derived rules and for pruning the worthless rules. In order to derive both Association rule and Negative Association rule effectively, this condition is indispensable.

Let $\{a,b\} \in RP$ be a fixed rare item-pair and let M be a mediator for the pair. By the definition of the Indirect Association rule, a rule $(M; \{\beta_a, \beta_b\})$ is an Indirect Association rule if and only if

$$\mu(M \Rightarrow \beta_x) \geq t_\mu, \quad \Longleftrightarrow \quad P(M) \leq t_{\mu,\beta_x} \times P(M \wedge \beta_x), \; x \in \{a,b\},$$

where

$$t_{\mu,\beta_x} = \frac{1 - t_\mu P(\neg\beta_x)}{P(\beta_x)}.$$

Theorem 1. *Let M_1 and M_2 be mediators for a rare item-pair $\{a,b\}$ such that $M_1 \subseteq M_2$. (1) If $P(M_2) > t_{\mu,\beta_x} \times P(M_1 \wedge \beta_x)$, then $\mu(M \Rightarrow \beta_x) < t_\mu$ for any M that satisfies $M_1 \subseteq M \subseteq M_2$.*
(2) If $P(M_1) \leq t_{\mu,\beta_x} \times P(M_2 \wedge \beta_x)$, then $\mu(M \Rightarrow \beta_x) \geq t_\mu$ for any M that satisfies $M_1 \subseteq M \subseteq M_2$.

The proof is clear since $P(M) \geq P(M_1)$ and $P(M \wedge \beta_x) \leq P(M_2 \wedge \beta_x)$. □

Corollary 1. *Let M be a mediator for a rare item-pair $\{a,b\}$ and let $c \in M$. If $P(M) > t_{\mu,\beta_x} \times P(c \wedge \beta_x)$, then there is no Indirect Association rule via mediator including the item c.*

3 Measure μ

An item-set X has ***support*** s, denoted as $sup(X)$, when $s \times 100\%$ of customer transactions D contain the items in X:

$$sup(X) = |F(X)|/|D|,$$

where $F(X) = \{T \in D \mid T \text{ includes } X\}$ and $|.|$ denotes the cardinality of a set. An association rule is an implication form of $X \Rightarrow Y$, where X and Y are item-sets satisfying $X \cap Y = \phi$. The support of $X \Rightarrow Y$ is the support of $X \cup Y$, that is, $sup(X \cup Y)$. The famous measure *confidence* for $X \Rightarrow Y$, denoted as $conf(X \Rightarrow Y)$, is the ratio, $sup(X \cup Y)/sup(X)$ and is often used for capturing the interestingness of the rule. When the number of transactions in database D is infinitely large, *sup* and *conf* can be interpreted according to probability theory as:

$$sup(X \cup Y) = P(X \wedge Y), \qquad conf(X \Rightarrow Y) = P(Y \mid X).$$

On the other hand, a rule of the form $X \Rightarrow \neg Y$ is called a Negative Association rule, and represents a rule that customers who buy an item-set X are not likely to buy at least one item in an item-set Y. The support and confidence for a Negative rule $X \Rightarrow \neg Y$ are defined as follow [4]:

$$sup(X \Rightarrow \neg Y) = P(X) - P(X \wedge Y), \quad conf(X \Rightarrow \neg Y) = P(\neg Y \mid X) = 1 - P(Y \mid X).$$

The measure confidence has, however, the defect when there is a Negative correlation between the item-sets. Brin [4] proposed the measure chi-squared test for capturing the correlation, especially Negative correlation, between the item-sets to reinforce a weakness of the measure confidence. It is not good enough because the value of chi-squared test depends on the number of transaction. Therefore too many worthless rules could be derived from large database such as POS data. We define a measure μ for $X \Rightarrow \beta_Y$ as follows:

$$\mu(X \Rightarrow \beta_Y) = \frac{P(X \mid \beta_Y) - P(X \mid \neg\beta_Y)}{P(X \mid \beta_Y)} = \frac{P(X \wedge \beta_Y) - P(X)P(\beta_Y)}{P(X \wedge \beta_Y)(1 - P(\beta_Y))}.$$

The numerator $P(X \mid \beta_Y) - P(X \mid \neg\beta_Y)$ denotes conditional measure of relationship of X associated with β_Y, and the denominator $P(X \mid \beta_Y)$ makes it normalized. Thus clearly $-\frac{P(\beta_Y)}{1-P(\beta_Y)} \leq \mu(X \Rightarrow \beta_Y) \leq 1$ holds. Moreover, we have

$$\mu(X \Rightarrow \beta_Y) = 0 \iff \phi(X, \beta_Y) = 0, \quad \mu(X \Rightarrow \beta_Y) > 0 \iff \phi(X, \beta_Y) > 0,$$

where ϕ is a correlation function, that is,

$$\phi(X,Y) = \frac{P(X \wedge Y) - P(X)P(Y)}{\sqrt{P(X)P(\neg X)P(Y)P(\neg Y)}}.$$

Zhang [15] presented new association rules for a measure of relationships between item-sets. He introduced two measures P_A and P_D to describe association and disassociation relationships between X and Y defined as follows:

$$P_A(X \Rightarrow Y) = \frac{P(X \wedge Y) - P(X)P(Y)}{P(X \wedge Y)(1 - P(Y))}, \text{ if } \phi(X,Y) \geq 0,$$
$$P_D(X \Rightarrow Y) = \frac{P(X)P(Y) - P(X \wedge Y)}{(P(X) - P(X \wedge Y))P(Y)}, \text{otherwise.}$$

The measure μ defined above is essentially based on measures P_A and P_D due to Zhang as seen in the following relations:

$$P_A(X \Rightarrow Y) = \mu(X \Rightarrow Y), \text{ if } \phi(X,Y) \geq 0,$$
$$P_D(X \Rightarrow Y) = -\mu(X \Rightarrow \neg Y), \text{ otherwise.}$$

Our measure μ is identical with the measure P_A if there is nonnegative correlation between X and Y. Otherwise, μ and P_D have opposite signs although the absolute values are equal. In this paper, we adopt the measure μ to measure both positive and Negative Association rules instead of P_A and P_D. Of course, we are concerned with an association rule $X \Rightarrow \beta_Y$ with positive correlation for X and β_Y, and so confine ourselves to rules $X \Rightarrow \beta_Y$ with $\mu(X \Rightarrow \beta_Y) \geq 0$. If a rule $X \Rightarrow \neg Y$ is a valid rule with the measure μ, this indicates that there is a Negative correlation between the item-sets X and Y. The set $\neg Y$ is equivalent to the set $X - X \cup Y$. In this paper, we are going to apply the measure μ to Indirect Association rule mining by judging from the point of view of properties for a good measure. By the definition of μ, it immediately follows that:

Theorem 2. *The measure μ satisfies the following five conditions:*

(1) $\mu(X \Rightarrow \beta_Y) = 0$ if X and Y are statistically independent, i.e., $P(X \wedge \beta_Y) = P(X)P(\beta_Y)$.

(2) $\mu(X \Rightarrow \beta_Y)$ monotonically increase with $P(X \wedge \beta_Y)$ when $P(X)$ and $P(\beta_Y)$ remain the same.

(3) $\mu(X \Rightarrow \beta_Y)$ monotonically decrease with $P(X)$ (or $P(Y)$) when the rest of parameters remain unchanged.

(4) $0 \leq \mu(X \Rightarrow \beta_Y) \leq 1$ if a correlation of X and β_Y is nonnegative.

(5) $\mu(X \Rightarrow \beta_Y) = 1$ if $P(X \wedge \beta_Y) = P(X)$.

The assertions (1), (2) and (3) for $\beta_Y = Y$ are proposed by Piatesky-Shapiro for a good measure for association rule $X \Rightarrow Y$. By the assertions (4) and (5), values of μ attain one if and only if X completely associates with β_Y. We convinced that a good measure μ should satisfy the above five conditions to understand the interestingness of the rules easily. For example, chi-squared test does not satisfy the conditions (1) and (2). It is widely known that the value of the chi-squared test depends on the number of transactions. Consequently too many worthless rules could be derived from a large database such as POS data.

3.1 Comparisons with the other measures

The measure μ is an efficient measure because it reflects the correlation and confidence. The relations between μ and confidence or correlation are as follows:

Theorem 3. *The following two properties hold:*
(1) $\mu(X \Rightarrow \beta_Y)$ monotonically increases with $conf(X \Rightarrow \beta_Y)$ when $P(\beta_Y)$ remains unchanged.
(2) $\mu(X \Rightarrow \beta_Y)$ monotonically decreases with $P(\beta_Y)$ when $conf(X \Rightarrow Y)$ remains unchanged.

Proof. We omit this proof for space constraint.

Secondly, we investigate a relation between our measure μ and IS measure. IS measure has been proposed in [9–12] for the first model of Indirect Association rule mining to measure dependencies between the mediator and the rare item-pair.They assert that IS measure is efficient metric because it contains both statistical dependence and statistical significance as follows:

$$IS(X,Y) = \frac{P(X \wedge Y)}{\sqrt{P(X)P(Y)}} = \sqrt{I(X,Y) \times sup(X \cup Y)},$$

where $I(X,Y) = \frac{P(X \wedge Y)}{P(X)P(Y)}$ is the measure Interest proposed in [4]. The measure IS takes in $I(X,Y)$ for statistical dependence and $sup(X \cup Y)$ for statistical significance. IS is unqualified for deriving Negative Association rules between a mediator and a rare item-pair because the value of the IS measure for the rule $X \Rightarrow Y$ is $\sqrt{P(X)P(Y)}$ when correlation is 0. Because $\sqrt{P(X)P(Y)}$ is according to the each of derived rules, we could have problems for deriving Negative Association rules. The first case, if the threshold of IS measure is set too high above $\sqrt{P(X)P(Y)}$, we could not derive Negative Association rules at all. The second case, if the threshold of IS measure is set too low, we could derive too many worthless rules. In this case, there would be many statistically insignificant rules, that is, the item-sets X and Y might be nearly independent statistically. As you have seen, the properties of IS measure is unqualified to detect interesting Negative Association rules. Furthermore IS measure does not satisfy the conditions 2 and 3 of Theorem 1. As you have seen, Chi-squared test and IS that are proposed as good measures do not satisfy the all properties for a good measure. Therefore we are going to propose the measures μ to derive Indirect Association rules effectively.

4 Experiments

4.1 Algorithm

The algorithm of mining Indirect Association rules can be composed into three phases. (1) Generate the sets of frequent items I_f and rare item-pairs RP. (2) Generate the sets of Mediators $\mathcal{M}$ for each item-pair in RP. (3) Generate all Indirect Association rules and prune the redundant rules. In step (1), generate the sets of frequent items I_f the first. If $sup(a)$ is greater than the frequent item threshold t_f, then a is appended into I_f. The second, generate the sets of rare item-pairs RP, (a, b) is appended into RP if $sup(a, b)$ is less than the rare item-pair threshold t_r. In step (2), for each item-pair in RP, generate the set of Mediators. Apriori-gen proposed in [1] is used to generate the set of Mediators. In step (3), outputs the item-pair and mediator $(M; (\beta_a, \beta_b))$ if $\mu(M \Rightarrow \beta_a) \geq t_\mu$ and $\mu(M \Rightarrow \beta_b) \geq t_\mu$.

4.2 Retail data

We have performed some data sets of POS data distributed for the domestic workshop of DATA MINING OLYMPIC 2000 held in KYOTO. Our algorithm has been written in Visual C++ for Windows. This experiment were performed on a 2.6GHz pentium 4 with 1GB main memory. These data sets are POS data of drug stores from April 1999 to March 2000. There are 5169613 transactions with 8476 items in database of 100 shops. In this experiment, thresholds are defined as follows:

$$t_r = 0.000001, t_m = 0.0001(t_r \times 100), t_f = 0.001(t_r \times 1000), t_\mu = 0.5.$$

For space constraint, we are going to show two figures illustrating some of derived Indirect Association rules from POS data. The first figure shows that some of derived Indirect Association rules for a rare item-pair that we chose at random. These rare items are not competitive products in our consideration because it seems that there is not any relationship between lipstick and detergent at all. The items of mediators for the first two rules are the same kinds of products except maker. We can get great insights that antiphlogistics and vitamin drink are tend to buy together and with lipstick but not with detergent. However customers who bought vitamin drink and any other item tend to buy detergent together, but not lipstick. So antiphlogistics could be regarded as key item for marketing. The second figure shows that some of derived Indirect Association rules for a rare item-pair that we chose with arbitrariness in order to analyze competing situations. Many customers tend to buy a facial tissue and a toilet paper. OUJINEPIA has two brands of facial tissue, hoxy and nepia. However customers who bought a facial tissue of nepia tend to buy a toilet paper of other competing maker. Customers who bought a facial tissue of hoxy tend to buy a toilet paper of the same brand. Furthermore customers who did not buy a facial tissue always tend not to buy a toilet paper of DAIOPAPER at all. Without the domain knowledge, many useful Indirect Association rules that describing the situation of competing were derived. These interpretations of derived rules could not be the real behavioral characteristics because our model does not take in representative explanatory variables, such as price and promotion. In our consideration, these rules are regarded as reliable from the point of view of underlying behavioral pattern. We can say that the first two rules in Figure 2 illustrate a co-essential behavioral pattern because each of items in the rules belongs to the same item category. Even though our model does not take in some information, we deem that these interpretations could be considered adequate and proper. We convinced that our model enables the users to understand the behavioral pattern of customers and the competitor.

Mediator	a		b	
Vitamin drink (A) Antiphlogistics (C)	$\Rightarrow a$	0.86	$\Rightarrow \neg b$	0.70
Vitamin drink (B) Antiphlogistics (D)	$\Rightarrow a$	0.80	$\Rightarrow \neg b$	0.73
Vitamin drink (A) Any other item	$\Rightarrow \neg a$		$\Rightarrow b$	
Vitamin drink (B) Any other item	$\Rightarrow \neg a$		$\Rightarrow b$	

COMPANY

A: TAIHO Pharmaceutical Co., Ltd.
B: TAISYO Pharmaceutical Co., Ltd.
C: KOBAYASHI Pharmaceutical Co.,Ltd.
D: PIP-FUJIMOTO Pharmaceutical Co., Ltd.
E: ROHTO Pharmaceutical Co., Ltd.
F: P&G

rare item pair a: Lipstick (E) and b: Detergent (F)

Figure 2

Mediator	a		b	
Tissue paper (A) Deodorant (C)	$\Rightarrow a$	0.80	$\Rightarrow \neg b$	0.67
Tissue paper (B) Deodorant (C)	$\Rightarrow a$	0.75	$\Rightarrow \neg b$	0.73
Tissue paper (A) Deodorant (C)	$\Rightarrow \neg a$	1	$\Rightarrow b$	0.92
Deodorant (C) Any other item	$\Rightarrow \neg a$	1	$\Rightarrow b$	

COMPANY

A: OUJINEPIA
B: DAIO-PAPER
C: CRECIA
D: KOBAYASHI Pharmaceutical Co., Ltd.

rare item pair a: Toilet paper (D) and b: Toilet paper (A)

Figure 3

5 Conclusion

We have discussed our new framework of Indirect Association rule mining with μ measure. Indirect Association rules enable us to form many interpretations about the competitors, the relation of the brands, and customer behavior. It had been hard to analyze competitors and to pick out the valid Negative Association rules without having the domain knowledge. Though we found many valid rules from POS data, we took no thought of the price. As a future work, we are going to extend this model cover the price and other informations.

References

1. R. Agrawal, T. Imielinski and A. Swami: *Mining Association Rules between Sets of Items in Large Databases*, in Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., USA, May 26–28, 1993, pp. 207–216, 1993.
2. R. Agrawal and R. Srikant: *Fast Algorithm for Mining Association Rules*, in Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, 1994, pp. 487–499, 1994.
3. R.J. Bayardo Jr. and R. Agrawal: *Mining the Most Interesting Rules*, in Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 15–18, 1999, pp. 145–154, 1999.
4. S. Brin, R. Motwani, and C. Silvertein: *Beyond market baskets: Generalizing association rules to correlations*, in Proceedings of 1997 ACM-SIGMOD International Conference on Management of Data, Tucson, AZ, 1997.
5. S. Brin, R. Motwani, J.D. Ullman and S. Tsur: *Dynamic Itemset Counting and Implication Rules*, in Proceedings of the ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA, May 13–15, 1997, pp. 255–264, 1997.
6. A. Savasere, E. Omiecinski and S. Navathe: *Mining for Strong Negative Associations in a Large Database of Customer Transactions*, in Proceedings of the Fourteenth International Conference on Data Engineering, Orlando, Florida, USA, February 23–27, 1998, pp. 494–502, 1998.
7. R. Srikant and R. Agrawal: *Mining Generalized Association Rules*, in Proceedings of the 21st International Conference on Very Large Data Bases, Zurich, Switzerland, September 11–15, 1995, pp. 407–419, 1995.
8. R. Srikant, and Q. Vu and R. Agrawal: *Mining Association Rules with Item Constraints*, in Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, California, USA, August 14–17, 1997, pp. 67–73, 1997.
9. P.N. Tan and V. Kumar: *Interestingness Measures for Association Patterns: A perspective*, Technical Report # TR00-036, Department of Computer Science, University of Minnesota, 2000.
10. P.N. Tan, V. Kumar and J. Srivastava: *Indirect Association: Mining Higher Order Dependencies in Data*, in Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery (PKDD-2000), Lyon, France, September 13–16, 2000, Lecture Notes in Computer Science **1910**, pp. 632–637, 2000.
11. P.N. Tan and V. Kumar: *Mining Indirect Associations in Web Data*, in Proceedings of the Third International Workshop on Mining Web Log Data Across All Customers Touch Points, San Francisco, CA, USA, August 26, 2001, Revised Papers, Lecture Notes in Computer Science **2356**, pp. 145–166, 2001.
12. P.N. Tan, and V. Kumar and J. Srivastava: *Selecting the Right Interestingness Measure for Association Patterns*, in Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, July 23–26, 2002, pp. 32–41, 2002.
13. P.N. Tan: *Discovery of Indirect Association and Its Applications*, Dr. Thesis, UNIVERSITY OF MINNESOTA, July, 2002.
14. T. Washio, H. Matuura, and H. Motoda: *Mining Association Rules for Estimation and Prediction*, in Proceedings of the Second Pacific-Asia Conference on Research and Development in Knowledge Discovery and Data Mining (PAKDD-98), Melbourne, Australia, April 15–17, 1998, Lecture Notes in Computer Science **1394**, pp. 417–419, 1998.
15. T. Zhang: *Association Rules*, in Proceedings of the Fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications (PAKDD-2000), Kyoto, Japan, April 18–20, 2000, Lecture Notes in Computer Science **1805**, pp. 245–256, 2000.
16. C. Zhang and S. Zhang: "Association Rule Mining," Lecture Notes in Artificial Intelligence **2307**, Springer-Verlag, 2002.